BAYESIAN DECISION THEORY IN VARIABLES SAMPLING FOR QUALITY CONTROL

James R. Anderson

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

Bayesian Decision Theory in Variables Sampling for Quality Control

by

James R. Anderson

June 1977

Thesis Advisor: B. O. Shubert

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Bayesian Decision Theory in Variables Sampling for Quality Control		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis June 1977 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) James R. Anderson		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93940		12. REPORT DATE June 1977 13. NUMBER OF PAGES 51
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Naval Postgraduate School Monterey, CA 93940		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Bayesian decision method has several features which are desireable in the sampling inspection process for quality control. These features include: (1) comparison of the value of sample information in the decision process with the cost of obtaining the information; (2) basing decisions on their consequences to the decision maker; and (3) allowing the use of subjective information

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (continued)

in the decision process. In this paper the Bayesian decision procedure as it applies to variables sampling for quality control is examined. The basic method is developed for both simultaneous and sequential sampling and the modeling of decision consequences is discussed. Various models for the production process are provided and solutions for the generalized linear model obtained. Finally the incorporation of subjective information is discussed.

Bayesian Decision Theory in Variables Sampling for Quality Control

by

James R. Anderson
B.S.E.E., Iowa State University, 1969

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 1977

ABSTRACT

The Bayesian decision method has several features which are desireable in the sampling inspection process for quality control. These features include: (1) comparison of the value of sample information in the decision process with the cost of obtaining the information; (2) basing decisions on their consequences to the decision maker; and (3) allowing the use of subjective information in the decision process. In this paper the Bayesian decision procedure as it applies to variables sampling for quality control is examined. The basic method is developed for both simultaneous and sequential sampling and the modeling of decision consequences is discussed. Various models for the production process are provided and solutions for the generalized linear model obtained. Finally the incorporation of subjective information is discussed.

TABLE OF CONTENTS

I. INTRODUCTION

One of the primary concerns in the quality control of items produced by or received from a production line are the procedures by which decisions are made concerning the quality of the material produced. In order to develop a decision procedure the abstract term "quality" must be operationally defined. This definition usually takes the form of an equipment specification which lists the characteristics required of the unit to perform its intended function. In the case of 100% inspection every unit produced or received is subjected to a test in which these characteristics are measured. Based on the test results the decision procedure is to accept those units which satisfy the specification and reject those which do not. When sampling inspection is used, a sample of the production output is tested. The test results of the sample are then used to make and accept or reject decisions concerning the population or lot from which the sample was drawn.

The decision procedure in this case requires that a decision function be specified which indicates, for the test results observed, the decision to be made (accept or reject). Unlike 100% inspection, in sample testing there always exists the possibility that the decision function may indicate an erroneous decision.

The consequences of erroneous decisions represent a loss to the decision maker and can range from mild to severe. As an example suppose a machine is judged to be out of calibration when in fact it is not. Then the loss to the decision maker would be the cost of a needless recalibration which may be small. On the other hand if production quality were judged to be acceptable when in fact it was not, consumers might seek alternate sources of supply. This could result in the loss of entire production contracts and reputation. Although the above examples are oversimplified the main idea is that erroneous decisions always represent a loss to the decision maker. Thus in the design of a decision procedure the loss due to erroneous decisions must be considered.

Another consideration in the design of a decision procedure is the cost of testing the sample units. This cost includes the labor and test facilities required and may include the cost of the units themselves (if the tests are destructive) or repair costs if the units fail. These costs may be large if complex facilities are required or test time is long. If the cost of testing is larger than the anticipated consequences of a decision then the cost of information is greater than its value in the decision process. Under these circumstances, gathering further information (testing) is counterproductive. Thus a decision procedure should indicate the "value' of additional information to the decision maker.

The most common method of specifying a decision procedure is based on classical hypothesis testing. In this approach two points on the operating characteristic (OC) curve for the decision procedure are specified. The OC curve for the procedure is the probability of acceptance versus equipment quality. The required sample size and reject/accept criteria are then developed based on the sampling distribution using a likelihood ratio test. Another method which provides greater flexibility and has features absent in the classical method is the Bayesian decision approach. In the Bayesian method the decision procedure is optimized for a loss function specified by the user which reflects this particular application. If the loss function and the cost of testing are expressed in the same units the cost of information can be obtained. The Bayesian method also contains the classical procedure as a special case. Another feature of the Bayesian method is that specific knowledge of the behavior of the production process as well as subjective information can be incorporated thus allowing the decision procedure to adapt to changing requirements.

In the following sections the basic Bayesian decision procedure will be outlined and the specification of loss functions and models for the production process discussed. The generalized linear model is introduced and the recursion equations developed to facilitate calculation of posterior distributions. Finally, the incorporation of subjective information is discussed.

II. THE BAYESIAN METHOD

A. THE GENERALIZED BAYESIAN DECISION PROCEDURE

In order to discuss the Bayesian method as applied to a production line, a generalized model of the production and sample test process is required. Let θ represent the characteristic of the equipments upon which decisions are to be made. For example θ would be the average or mean gain of a production lot of amplifiers. The actual value of θ is not observable, however, we can perform tests which indicate the gain of an individual amplifier. Let x indicate the results of such test. Also let θ and x be related through a known probability density function denoted by $f(x|\theta)$. As a model of the generalized production process we assume a random process such that for each time t, θ_τ has continuous distribution. It is also assumed that there exists a time increment $\Delta t>0$ for which θ is constant i.e., $\theta_t = \theta_{t+\Delta\tau}$ for all t. This assumption implies that given a production increment of n items produced during Δt, test results for each unit are samples from $f(x|\theta_t)$. Figure 1 shows how θ might vary for the generalized process. As shown in the figure, θ for the increment Δt is a fixed but unknown quantity. It is the units produced during each Δt which are the object of the decision process. At each end point t_1, t_2, ... a decision must be made to either accept or reject the units produced in the

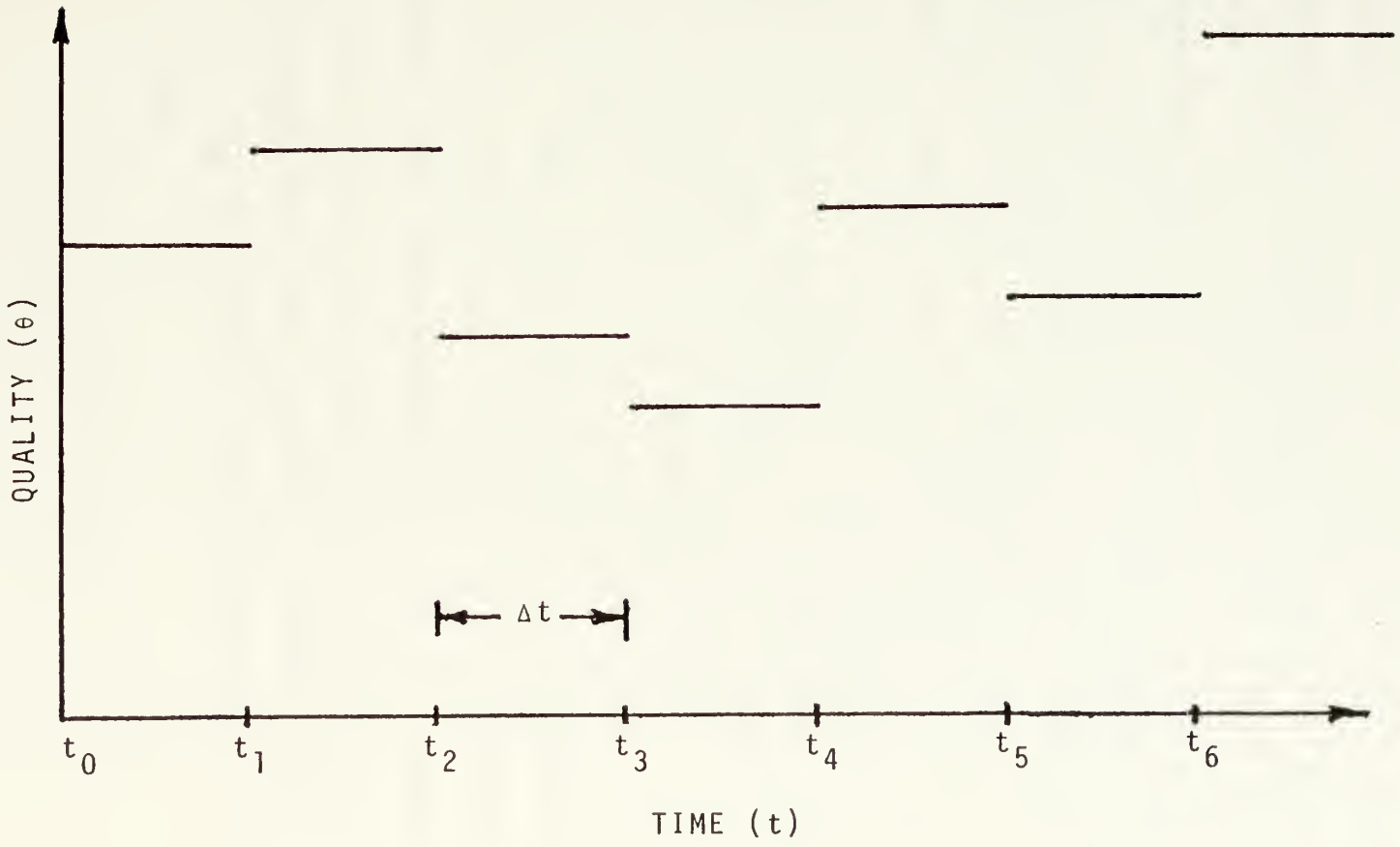

Figure 1. Generalized Production Process.

preceding time increment. The decisions will be based on the sample data (x) which provides information concerning the true value of θ. At time t_0 let the probability density $f(\theta_o)$ represent the uncertainty as to the true value of θ_o. Since the value of θ at time t may or may not be independent of the previous values of $\theta(\theta_{t-1}, \theta_{t-2} \ldots)$ let the uncertainty of θ_t given $(\theta_{t-1}, \theta_{t-2} \ldots)$ be expressed by means of the conditional density $f(\theta_t|\theta_{t-1}, \theta_{t-2} \ldots)$ which is known or can be estimated but not necessarily the same for all t. Since the decisions (accept or reject) are to be based on the characteristic θ, it is necessary to specify a loss function which indicates the consequences of a particular decision when a specific value of θ obtains, for all possible values of θ and all decisions. Let Θ represent the set of all possible values of θ and let D represent the set of all decisions d. Then let $L(\theta,d)$ be the loss function which is known for all $d \varepsilon D$ and $\theta \varepsilon \Theta$. The loss for a particular decision $d \varepsilon D$ depends on the actual value of θ, but θ is unknown. The expected loss of decision d would be the product of $L(\theta,d)$ times the probability that θ obtains, summed over all possible values of θ. Let $\rho(d)$ denote the expected loss or risk of decision d then:

$$\rho(d) = \int_{\Theta} L(\theta,d)\, f(\theta) d\theta \quad . \qquad (1)$$

The optimal decision in terms of minimizing the risk is the decision $d \epsilon D$ which minimizes equation (1) and is denoted by d^* therefore:

$$\rho(d^*) = \min_{d \epsilon D} \int_{\Theta} L(\theta,d)\ f(\theta)d\theta \ . \qquad (2)$$

Thus given a loss function $L(\theta,d)$ and the distribution of θ, $f(\theta)$, the optimal decision is defined by equation (2). The optimal decision (d^*) is usually referred to as the Bayes decision against $f(\theta)$. Since the decisions (accept or reject) will be based on the sample data ($\underline{x}:[x_1,x_2, \ldots, x_n]$) it is desirable to specify a decision function, denoted by $\delta(\underline{x})$, which for every value of $\underline{x}$ observed, specifies the decision d^*, i.e., the decision which minimizes the risk. Let S_n be the set of all possible test results. Then the risk of the decision function $\delta(\underline{x})$ is by equation (1)

$$\rho(\delta(\underline{x})) = \int_{\Theta} \int_{S_n} L(\theta,\delta(\underline{x}))\ f(\underline{x}|\theta)\ f(\theta)d\underline{x}d\theta. \qquad (3)$$

We want to find the decision function $\delta^*(\underline{x})$ which minimized the risk as expressed in equation 3. Assuming $L(\theta,\delta(\underline{x}))$ is bounded, interchanging the order of integration yields:

$$\rho(\delta(x)) = \int_{S_n} \left[\int_{\Theta} L(\theta,\delta(\underline{x}))f(\underline{x}|\theta)f(\theta)d\theta \right] d\underline{x} \ . \qquad (4)$$

Equation (4) is minimized when the expression in brackets is minimized for each value of $\underline{x} \epsilon S_n$. Thus the optimal decision function $\delta^*(\underline{x})$ would specify a decision d^* which minimized the integral

$$\int_{\Theta} L(\theta,d)f(\underline{x}|\theta)f(\theta)d\theta \ . \tag{5}$$

From Bayes theorem $f(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)f(\theta)}{f(\underline{x})}$ (6)

where $f(\underline{x}) = \int_{\Theta} f(\underline{x}|\theta)f(\theta)d\theta$ is a constant given a sample value for $\underline{x}$. Then minimizing the integral of (5) is equivalent to selecting a decision d* which minimizes:

$$\int_{\Theta} L(\theta,d) \frac{f(\underline{x}|\theta)f(\theta)}{f(\underline{x})} d\theta = \int_{\Theta} L(\theta,d)f(\theta|\underline{x})d\theta \tag{7}$$

for each value of $\underline{x}$ observed. Thus it is not necessary to determine in advance a decision function $\delta(\underline{x})$ which specifies d* for all possible values of $\underline{x}$. As each $\underline{x}$ is observed the posterier, $f(\theta|\underline{x})$, is calculated from the prior, $f(\theta)$, by equation (6) and d* is chosen to satisfy

$$\rho(d^*) = \min_{d\varepsilon D} \int_{\Theta} L(\theta,d)f(\theta|\underline{x})d\theta \tag{8}$$

This is the same result as equation (2) except that the posterior based on the test data $\underline{x}$ is used instead of the prior. Thus equation (2) defines the optimal decision d* before and after sampling as long as the appropriate value for $f(\theta)$ is used.

The next step is to determine whether the decision should be made without sampling based on the prior distribution, $f(\theta)$, or the sample tested and the decision based on the posterior distribution $f(\theta|\underline{x})$. In order to determine which action is optimal the risk of obtaining an additional sample

and then proceeding in an optimal fashion must be obtained. If the risk of selecting a decision immediately is greater than the risk of obtaining a sample result and then proceeding in an optimal manner, then the sample should be tested since this is the minimum risk action. Let $\rho(\phi,\underline{x})$ denote the risk of obtaining a sample $\underline{x}$ when the prior of θ is ϕ and then proceeding in an optimal manner. Also let $\rho(\phi,d^*)$ denote the risk of making decision d^* when the prior of θ is ϕ. Then the following decision rule will be used.

If $\rho(\phi,d^*) > \rho(\phi,\underline{x})$ test the sample; (9)

otherwise, make decision d^*. $\rho(\phi,d^*)$ is obtained from equation (2) where $\phi = f(\theta)$ is:

$$\rho(\phi,d^*) = \min_{d\varepsilon D} \int_{\Theta} L(\theta,d)f(\theta)d\theta \ . \tag{10}$$

To determine $\rho(\phi,\underline{x})$ two cases must be distinguished; the samples are tested simutaneously or sequential sampling is used.

1. Simultaneous Sampling

It has been assumed that the cost of obtaining sample results is not zero. Therefore let $C_n(\underline{x})$ represent the cost of testing the units 1, 2, ..., n where test results are represented by the vector $\underline{x} = (x_1, x_2, \ldots, x_n)$. In many cases the cost of testing is independent of the values obtained, in which case the cost would be just C_n, the cost of testing n units. The expected loss of testing n units and then making the optimal decision d^* plus the cost of testing

is $\rho(\phi,\underline{x})$. The distribution of θ if $\underline{x}$ is observed will be $f(\theta|\underline{x})$ thus from equation (10):

$$\rho(\phi,\underline{x}) = \int_{S_n} \left[\min_{d\varepsilon D} \int_{\Theta} L(\theta,d) f(\theta|\underline{x}) d\theta \right] f(\underline{x}) d\underline{x} + E[C_n(\underline{x})] \quad (11)$$

Thus the decision procedure for the production lot would be as follows:

1. Prior to testing determine $\rho(\phi,d^*)$ from equation (10) based on the prior $f(\theta)$.
2. Determine the risk of testing, $\rho(\phi,\underline{x})$, from equation (11).
3. If $\rho(\phi,d^*) < \rho(\phi,\underline{x})$ make decision d^* otherwise.
4. Test sample units to obtain data $\underline{x}$ and make decision d^* according to equation (8).

2. Sequential Sampling

The risk of sampling in a sequential procedure differs from the simultaneous case because after a sample is tested two actions are possible, (a) make a decision or (b) continue sampling. Because of this, determining the risk of a sequential procedure is, in general, more difficult than determining the risk of the simultaneous case just discussed. In most cases of practical interest the sample size has a fixed upper bound. Let n denote the maximum number of samples available for testing. Then the risk of testing the first unit and proceeding in an optimal fashion is the risk of the n step sampling procedure where after each sample is tested

the risk of continuing is compared with the risk of choosing a decision and the minimum of the two risks is chosen. After the first sample is taken, the risk of choosing a decision must be compared to the risk of continuing with the n-1 step sampling procedure. Thus as each sample is drawn the risk of continuing changes due to the change in the sample number remaining as well as the new prior based on the samples observed. The above process may be viewed as a decision tree shown in figure 2 which depicts the sequential decision process for n=4. At each step (k) k = 0, 1, 2, ..., n the risk of making an immediate decision is denoted by $\rho(\phi_k, d^*)$ where ϕ_k is the distribution of θ based on k samples and is defined in equation (10). At each step this must be compared with the risk of continuing the sequential test process and the minimum risk action chosen. The risk of continuing at each step is denoted by $\rho(\phi_k, \underline{x}_{k+1})$ where $\underline{x}_{k+1} = [x_{k+1}, x_{k+2}, \ldots, x_{n-1}, x_n]$ indicating the dependence on the current prior and the remaining samples.

The difficulty alluded to earlier is in obtaining values for $\rho(\phi_k, \underline{x}_{k+1})$. The general solution procedure uses a backward induction starting at the last step and working backward to the first step to obtain the continuation risk at each step. For the n=4 case depicted in Figure 2 the procedure would be as follows. At step 3 after three samples had been observed the optimal action would be the minimum of $\rho(\phi_3, d^*)$ and $\rho(\phi_3, \underline{x}_4)$. Where $\rho(\phi_3, d^*)$ is the risk of the optimal decision given ϕ_3 as defined in equation (10) and

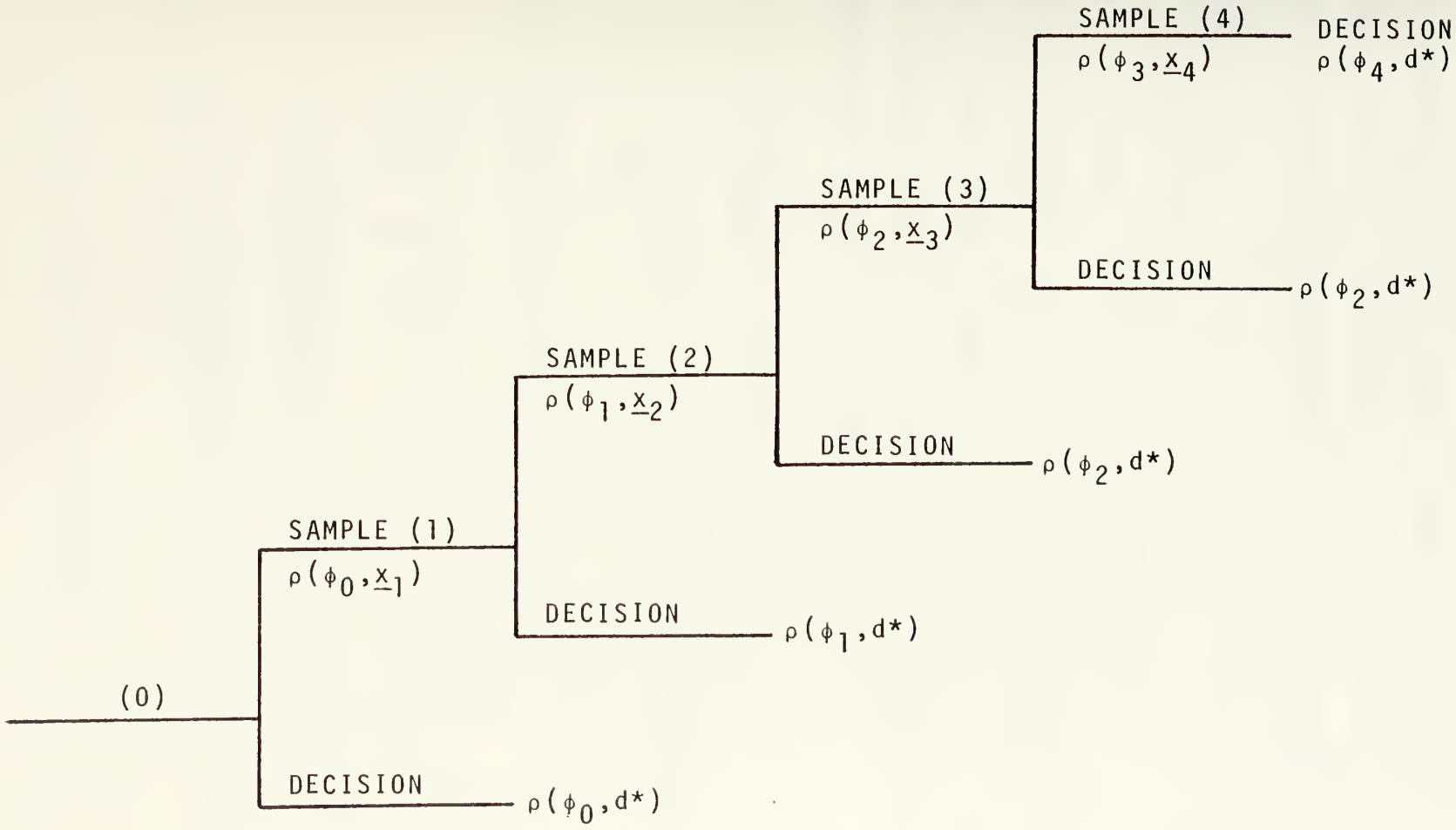

Figure 2. Sequential Decision Process.

$\rho(\phi_3, \underline{x}_4)$ is the risk of obtaining an additional sample x_4 and then making the optimal decision. Since a decision must be made after x_4 is observed this is the same as the sampling risk for simultaneous sampling defined in equation (11) with $\underline{x}$ equal to x_4. Thus the risk at step 3 is a function of ϕ_3 denoted by $\rho(\phi_3)$ and would be:

$$\rho(\phi_3) = \min [\rho(\phi_3, d^*), \rho(\phi_3, \underline{x}_4)]. \tag{12}$$

The risk at step 2 will be the minimum of the decision risk $\rho(\phi_2, d^*)$ and the continuation risk $\rho(\phi_2, \underline{x}_3)$. The continuation risk $\rho(\phi_2, \underline{x}_3)$ is the expected value of the risk at step 3 based on the sample, x_3. Thus,

$$\rho(\phi_2, \underline{x}_3) = E[\rho(\phi_3(x_3))] = \tag{13}$$

$$\int_{X_3} \min[\rho(\phi_3(x_3), d^*), \rho(\phi_3(x_3), \underline{x}_4)] f(x_3) dx_3 + C_3$$

where: $\phi_3(x_3)$ is ϕ_2 given x_3 i.e. $f_2(\theta | x_3)$

$$f(x_3) = \int_{\Theta} f(x_3|\theta) f_2(\theta) d\theta$$

C_3 = expected cost of obtaining value x_3

Thus at step 2, the optimal action again being that with minimal risk, the risk is

$$\rho(\phi_2) = \min[\rho(\phi_2, d^*), \rho(\phi_2, \underline{x}_3)]. \tag{14}$$

This approach is then repeated for step 1 which yields

$$\rho(\phi_1, \underline{x}_2) = E[\rho(\phi_2(x_2))] =$$

$$= \int_{x_2} \min [\rho(\phi_2(x_2), d^*), \rho(\phi_2(x_2)] f(x_2) dx_2 + C_2 \tag{15}$$

and

$$\rho(\phi_1) = \min[\rho(\phi_1, d^*), \rho(\phi_1, \underline{x}_2)] \ . \tag{16}$$

Thus at step 0, the beginning of the procedure, the risk of the entire sequential test procedure would be

$$\rho(\phi_0) = \min[\rho(\phi_0, d^*), \rho(\phi_0, \underline{x}_1)] \tag{17}$$

where $\rho(\phi_0, \underline{x}_1) = E[\rho(\phi_1(x_1))]$ is the risk of the entire sequential sampling plan.

As seen from the above discussion determining the risk of a sequential procedure is a non-trivial exercise. The degree of difficulty depending on the sample size n, the loss function $L(\theta, d)$ and the sampling and parameter distributions, $f(x|\theta)$ and $f(\theta)$. Examples of the above procedure for sequential testing may be found in the open literature [1, 2 and 3].

3. Comparison of Sequential Versus Simultaneous Sampling

In the design of a quality control procedure the method of sampling must be specified. In order to determine which of the two methods is preferred in a given situation the risks of the two procedures should be compared. In lieu of mitigating circumstances such as ease of implementation,

increased complexity, etc., the decision as to which sampling scheme, sequential or simultaneous, is optimal should be based on the risks of the two procedures. That is if $\rho(s^*)$ is the risk of the optimal sampling scheme s^* then

$$\rho(s^*) = \min[\rho(s_Q), \rho(s_i)] \ . \tag{18}$$

where s_Q = sequential sampling

s_i = simultaneous sampling

In many cases the decision as to which sampling scheme is superior is obvious due to the nature of the test. For example, if the cost of testing is constant regardless of the number of units tested then simultaneous testing would provide minimum risk. If however the cost of testing were only a function of the number of units tested then sequential testing would be superior. It is when the cost of testing assumes some combination of the two extremes that the optimal choice becomes unclear, in which case the risks of each procedure must be compared to determine the optimal approach.

B. APPLICATION TO THE GENERALIZED PRODUCTION PROCESS

In order to gain insight to the use and requirements of the Bayesian decision method, the implementation of the procedure on the generalized process of Figure 1 will now be discussed for the simultaneous sampling case.

At time t_0, t_1, t_2, ... the following assumptions are made.

1. A production sample of size n_t from production lot Q_t is available for testing.
2. Samples are independent given θ_t and the sampling density $f(x|\theta_t)$ is known.
3. The cost of testing the n units, C_n, is known.
4. A loss function $L(\theta,d)$ is specified for all $d\varepsilon D$ and $\theta\varepsilon\Theta$.
5. At time t_0, prior to sampling, the distribution of θ is known and denoted by $f(\theta_0)$.
6. The conditional distribution of θ_{t+1} given θ_t, θ_{t-1}, ... is known and denoted by $f(\theta_{t+1}|\theta_t)$ where a markov dependence is assumed for illustration.

At time t_0, prior to sampling, two actions are feasible. Either make a decision (accept or reject) or test the sample to gain information. If a decision is made without sampling the risk will be $\rho(d^*)$ as defined by equation (8). The risk of testing the sample $\underline{x} = (x_1, x_2, \ldots, x_n)$ and making the optimal decision, $\rho(\phi,\underline{x})$ is given by equation (11). Assume that sampling represents the minimal risk action. After the sample result is obtained, the prior $f(\theta_0)$ must be revised and the optimal decision d* chosen. Denote the posterior or new prior based on the data sample by $f(\theta_0|\underline{x})$, then by Bayes theorem

$$f(\theta_0|\underline{x}) = \frac{f(\underline{x}|\theta_0)f(\theta_0)}{f(\underline{x})} \quad .$$

The posterior $f(\theta_0|\underline{x})$ is now used to determine the optimal decision d*, by equation (10)

$$\rho(d^*) = \min_{d\in D} \int_{\Theta} L(\theta,d)\ f(\theta_0|\underline{x})\ d\theta_0.$$

After the decision is made on lot Q_0 at t_0 the procedure steps to lot Q_1 at t_1. In order to determine the appropriate actions concerning this lot the distribution $f(\theta_1)$ must be obtained. It is at this point where the model of the production process is used. The relationship between θ_0 and θ_1 must be known in order to determine the density $f(\theta_1)$ based on the posterier $f(\theta_0|\underline{x})$. The relationship between θ_0 and θ_1 is specified by the conditional density $f(\theta_1|\theta_0)$ which is obtained from the model of the production process. Methods by which this density may be obtained from the production model are discussed in a later section. Given that $f(\theta_1|\theta_0)$ is known then $f(\theta_1)$ prior to sampling from lot Q_1 is obtained as follows:

$$f(\theta_1) = \int_{\Theta} f(\theta_1|\theta_0)\ f(\theta_0|\underline{x})d\theta\ . \qquad (19)$$

Using this value as the prior for θ_1; $\rho(d^*)$ and $\rho(\phi,\underline{x})$ are obtained using equations (10) and (11) as before and the decision rule (9) is applied. If the decision rule indicates that the risk can be lowered by sampling, the procedure as outlined for θ_0 is followed. If however, no sampling is required then decision d* is made and the procedure advances

to t_2. At t_2, $f(\theta_2)$ must be obtained based on $f(\theta_0|\underline{x})$ since no samples were observed at time t_1. The density $f(\theta_2)$ would be determined as follows:

$$f(\theta_2) = \int_{\Theta} f(\theta_2|\theta_0)\ f(\theta_0|\underline{x})\ d\theta_0 \qquad (20)$$

where $f(\theta_2|\theta_0) = \int_{\Theta} f(\theta_2|\theta_1)\ f(\theta_1|\theta_0)\ d\theta_1$.

After $f(\theta_2)$ is determined $\rho(d^*)$ and $\rho(\phi,\underline{x})$ are obtained as before and the decision rule applied. The entire procedure is then repeated to determine if samples should be tested at t_3, t_4, ..., etc.

Under the decision process described it may be possible that no sampling would be required for several production lots. At first thought this may seem contrary to the objective of minimizing the decision risk. If the production sequence θ_0, θ_1, θ_2 ... is highly correlated then knowledge of one value of θ implies considerable knowledge of succeeding (and preceding) values. The correlation is expressed by the density $f(\theta_t|\theta_{t-1})$ which is derived from the model of the production process. The decision process thus quantifies the feeling "When one lot is good the next one usually is good also."

In order to apply the Bayesian method in loss function, $L(\theta,d)$, and the sampling and process densities, $f(x|\theta)$ and $f(\theta_t|\theta_{t-1})$ must be specified. These are the subject of the following sections.

III. LOSS FUNCTIONS

The purpose of this section is to examine various ways in which the consequences of decisions can be related to the true value of quality. In the preceding section this relationship was generally referred to as a loss function, $L(\theta,d)$. As mentioned previously the loss when the best decision is made for a given value of θ is equal to zero. The loss of a particular decision d when $\Theta = \theta$ is the difference between the consequences if d is chosen and the consequences if the best decision were chosen. The loss then essentially represents a regret or opportunity cost. From the above definition it is seen that one characteristic of loss functions is that they are non-negative. Since in the decision process the risk of sampling is added to the cost of testing the loss function and testing cost must be expressed in similar units (e.g., dollars). In the following examples it is assumed that the utility of money is linear over the range of interest. This assumption alleviates the otherwise necessary transformation of the loss in dollars to utility. If the utility of money is continuous then at least to a first order approximation the linear assumption is valid. In the following paragraphs several examples of loss functions are discussed. Their presences is not meant to imply that they are in any way the best or most useful loss functions. The loss function used for a particular process depends entirely on the situation at hand.

A. SYMMETRIC LINEAR LOSS FUNCTION

The following example demonstrates that if the decision consequences are a linear function of θ then the loss function will be linear and symmetric about their intersection. Let $R(\theta,d_i)$ represent the consequences of decision d_i when θ obtains. For the linear case

$$R(\theta,d_1) = a_1\theta + b_1 \tag{21}$$

$$R(\theta,d_2) = a_2\theta + b_2 \quad , a_2 > a_1 \; . \tag{22}$$

Equations (21) and (22) are plotted in figure 3.

If the consequences are viewed as a cost then the best decision is that which minimizes $R(\theta,d)$ for all values of θ Thus for $\theta \leq \theta_0$ decision d_2 is best and d_1 is best for $\theta \geq \theta_0$. The loss $L(\theta,d_1)$ is

$$L(\theta,d_1) = \begin{cases} a_1\theta + b_1 - (a_2\theta + b_2) & \theta \leq \theta_0 \\ 0 & \theta > \theta_0 \end{cases}$$

or

$$L(\theta,d_1) = \begin{cases} \frac{(b_1-b_2)}{(a_2-a_1)}(a_2-a_1) - \theta(a_2-a_1) & \theta > \theta_0 \\ 0 & \theta \leq \theta_0 \end{cases}$$

From (21) and (22) $\theta_0 = \dfrac{(b_1-b_2)}{(a_2-a_1)}$

Thus

$$L(\theta,d_1) = \begin{cases} (a_2-a_1)(\theta_0-\theta) & \theta < \theta_0 \\ 0 & \theta \geq \theta_0 \end{cases} \tag{23}$$

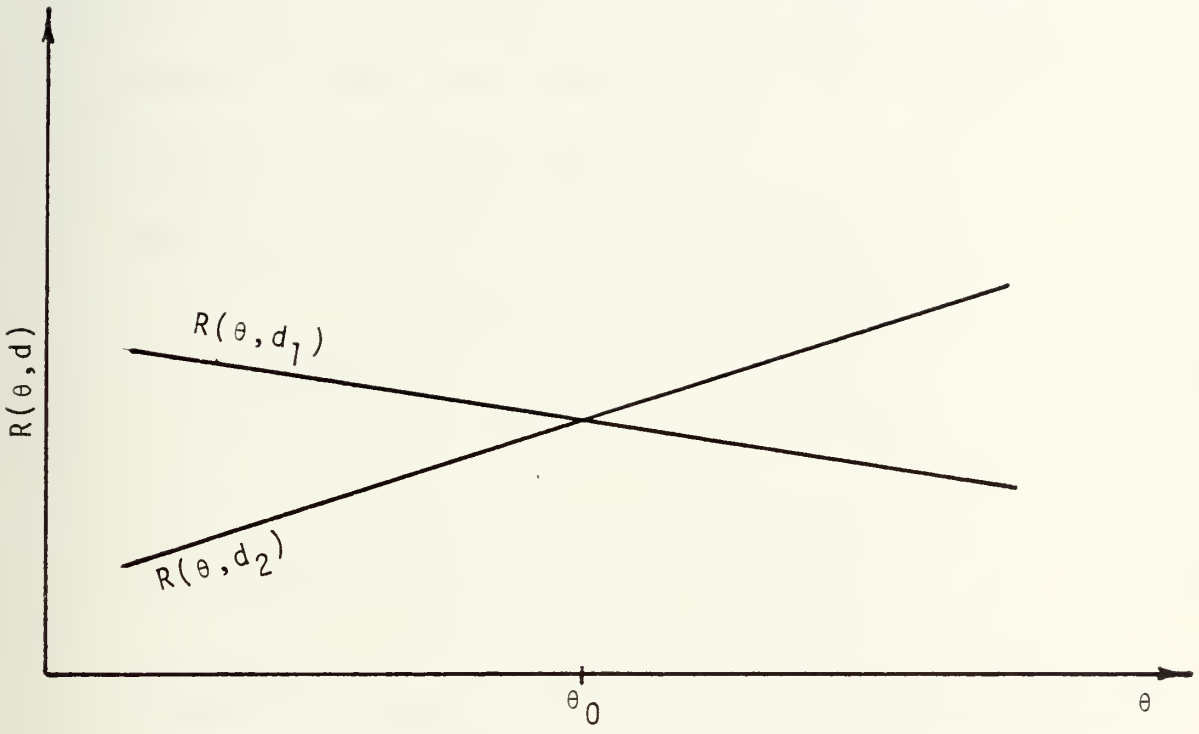

Figure 3. Decision Consequences.

By the similar calculations:

$$L(\theta,d_2) = \begin{cases} 0 & \theta \leq \theta_0 \\ (a_2-a_1)\,(\theta-\theta_0) & \theta > \theta_0 \end{cases} \tag{24}$$

Equations (23) and (24) are shown in figure 4

As evidenced by equations (23) and (24) and depicted in figure 4 the loss functions of linear consequences are linear and symmetric about the intersection.

B. QUADRATIC LOSS FUNCTIONS

The quadratic loss function is defined as:

$$L(\theta,d_1) = \begin{cases} C_1(\theta-\theta_0)^2 & \theta < \theta_0 \\ 0 & \theta \geq \theta_0 \end{cases} \tag{25}$$

$$L(\theta,d_2) = \begin{cases} 0 & \theta < \theta_0 \\ C_2(\theta_0-\theta)^2 & \theta > \theta_0 \end{cases} \tag{26}$$

A heuristic justification for this general form for a loss function can be made as follows. Assume for a particular problem that the loss function $L(\theta,d)$ and all its derivitives exist. Let θ_0 represent the dividing line between acceptable and unacceptable quality. If $\theta > \theta_0$ let d_1 be the proper decision and if $\theta < \theta_0$ let d_2 be the proper decision. Thus $L(\theta,d_1) = 0$ for $\theta > \theta_0$ and $L(\theta,d_2) = 0$ for $\theta < \theta_0$. Define $L(\theta) = \max_{d\in D} L(\theta,d)$ then $L(\theta) = L(\theta,d_1)$ for $\theta < \theta_0$ and $L(\theta) = L(\theta,d_2)$ for $\theta \geq \theta_0$. $L(\theta)$ can be expressed as a Taylor series expension about θ_0 as follows:

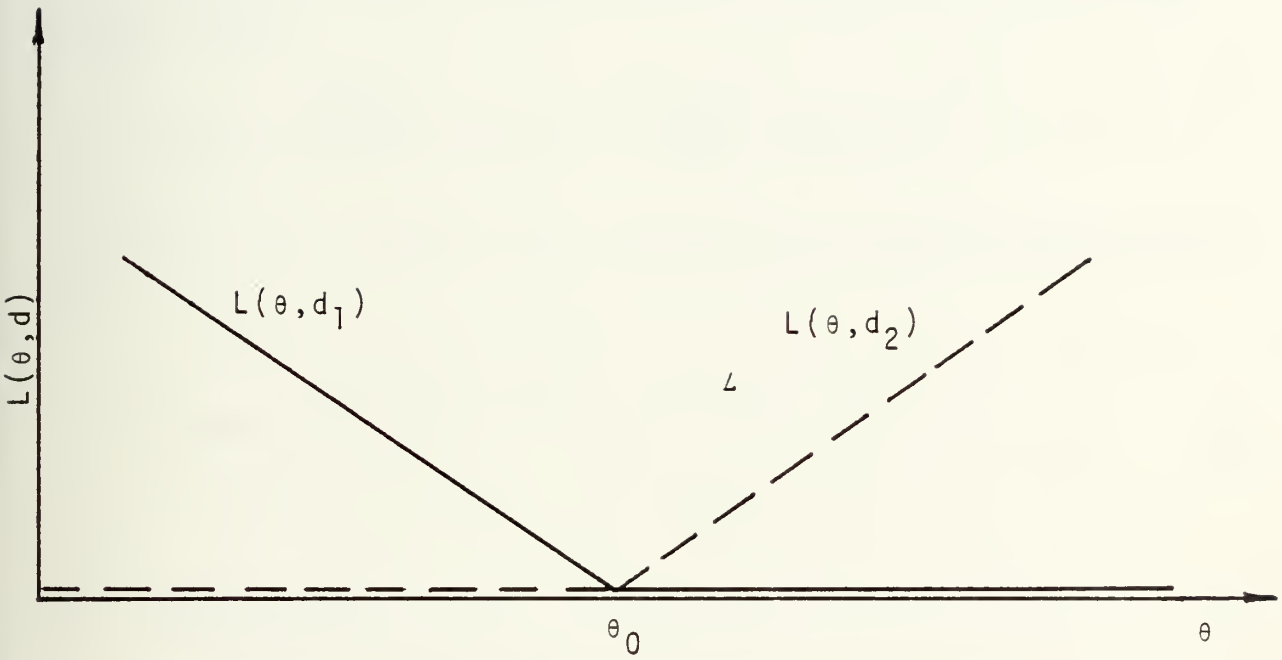

Figure 4. Linear Loss Functions.

$$L(\theta) = L(\theta_0) + L'(\theta_0)(\theta-\theta_0) + \frac{L''(\theta_0)}{2}(\theta-\theta_0)^2 + \ldots \quad (27)$$

If $\theta = \theta_0$ then either decision d_1 or d_2 is optimal and the loss is zero. This implies that $L(\theta_0) = 0$. If $L(\theta_0)$ is zero then θ_0 is a minimum for $L(\theta)$ which implies that $L'(\theta_0) = 0$. Further if $L(\theta_0)$ is a minimum then $L''(\theta_0)$ must be non-negative.

Applying these results to the Taylor series expension (27) yields:

$$L(\theta) = \frac{L''(\theta_0)}{2}(\theta-\theta_0)^2 + \frac{L'''(\theta_0)}{3!}(\theta-\theta_0)^3 + \ldots$$

Thus the loss function for θ close to θ_0 can be approximated by:

$$L(\theta,d_1) = \begin{cases} C_1(\theta-\theta_0)^2 & \theta < \theta_0 \\ 0 & \theta \geq \theta_0 \end{cases} \quad (28)$$

$$L(\theta,d_2) = \begin{cases} 0 & \theta \leq \theta_0 \\ C_2(\theta-\theta_0)^2 & \theta > \theta_0 \end{cases} \quad (29)$$

which is the quadratic loss function originally defined. Instead of specifying the indifference value θ_0 two values θ_1,θ_2 could be specified. Where θ_1 would represent minimum acceptable quality and θ_2 would represent the maximum rejectable quality. Then for d_1 = accept and d_2 = reject the loss function could be expressed as follows.

$$L(\theta,d_1) = \begin{cases} C_1(\theta-\theta_0)^2 & \theta \leq \theta_1 \\ 0 & \theta > \theta_1 \end{cases}$$

$$L(\theta,d_2) = \begin{cases} 0 & \theta \leq \theta_2 \\ C_2(\theta-\theta_0)^2 & \theta > \theta_2 \end{cases}$$

C. CONSTANT LOSS FUNCTION

The constant loss function represents the case where the loss is independent of the value of θ over a specified range. The constant loss function could be represented as follows:

$$L(\theta,d_1) = \begin{cases} 1 & \theta \leq \theta_0 \\ 0 & \theta > \theta_0 \end{cases} \tag{30}$$

$$L(\theta,d_2) = \begin{cases} 0 & \theta < \theta_0 \\ 1 & \theta \geq \theta_0 \end{cases} \tag{31}$$

The expected loss or risk, $\rho(d)$, for decisions d_1 and d_2 would be:

$$\rho(d_1) = \int_{-\infty}^{\theta_0} 1 \cdot f(\theta)d\theta = P(\theta < \theta_0) = \beta$$

$$\rho(d_2) = \int_{\theta_0}^{\infty} 1 \cdot f(\theta)d\theta = P(\theta > \theta_0) = \alpha$$

The risk for decision d_1 is the probability that d_2 is the correct decision and the risk for d_2 is the probability that d_1 is the correct decision. $\rho(d_1)$ and $\rho(d_2)$ are usually referred to as the probability of type II and type I errors respectively and denoted by β and α. It can be shown [1] that the decision function $\delta(x)$ which minimize the risk has the form. If $\frac{f_1(\underline{x})}{f_2(\underline{x})} < C$ then reject.

where

$$f_1(\underline{x}) = f(\underline{x}|\theta_0)$$

$$f_2(\underline{x}) = f(\underline{x}|\hat{\theta})$$

and $\hat{\theta}$ is the maximum likelihood estimate for θ based on $\underline{x}$.

This decision function is the generalized likelihood ratio criteria upon which classical hypothesis testing is based. Thus quality control procedures based on classical hypothesis testing imply that the loss functions of (30) and (31) are operative.

This concludes the discussion of the most common types of loss functions. The next section considers the problem of formulating a statistical model of the production process.

IV. THE PRODUCTION MODEL

As mentioned previously in order to apply the Bayesian decision method the conditional density $f(\theta_t|\theta_{t-1})$ must be known or estimated. In order to specify this density the quality control specialist must specify the underlying mechanisms which determine how the production process is evolving over time. In the following models two assumptions will be made: (a) that the process is determined by a specific underlying relationship plus random disturbances and (b) that by the central limit theorem, the ghost of LaPlace or some other incantation, the disturbances are assumed to be normally distributed with known mean and variance. In the following paragraphs three models which might be used to characterize a production process are described. They are the linear trend model, the autoregressive model and the periodic model. The generalized linear model is also presented. In addition to the process models a typical observation model will be included for completeness. For convenience and to aid in the later development of the solutions for the generalized linear model it is also assumed that the observation errors are normally distributed.

A. LINEAR TREND MODEL

The linear trend model represents a production process where the underlying trend is a linear increase or decrease in the characteristic θ. Let δ_τ represent the change of θ from t-1 to t and x_t represent the observations at time t. Then the observation and process models could be described as:

$$\text{OBSERVATION:} \quad x_t = \theta_t + e_t \qquad e_t \sim N(0,\sigma_{x_t}^2) \tag{32}$$

$$\text{PROCESS:} \quad \theta_t = \theta_{t-1} + \delta_t + w_t \qquad w_t \sim N(0,\sigma^2) \tag{33}$$

In the above model e_t represents observation noise and w_t represents the process noise causing deviations from the linear relationship. From the model, the conditional density $f(\theta_t|\theta_{t-1})$ would be normal with mean = $E(\theta_t) = \theta_{t-1} + \delta_t$ and variance equal to σ^2. In order to apply the Bayesian procedure the increment δ_t and the variance σ^2 must be known or estimated from the process. If the uncertainty in δ_t is incorporated in the model the result is

$$\theta_t = \theta_{t-1} + \delta_t + w_t \qquad w_t \sim N(0,\sigma^2)$$

$$\delta_t \sim N(\mu_R, \sigma_R^2)$$

then $f(\theta_t|\theta_{t-1}) \sim N(\theta_{t-1} + \mu_R, \sigma_R^2 + \sigma^2)$ assuming δ_t and w_t are uncorrelated.

B. AUTOREGRESSIVE MODEL

The basic autoregressive results from the following assumptions about the production process.

1. $f(\theta_t) \sim N(0,\sigma^2) \quad \forall t$

2. $f(\theta_t,\theta_{t+1}) \sim N(\underline{0},C) \quad \forall t,\ C = \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix}$

Assumption 1 indicates that at any time t the uncertainty in the location of θ_t can be expressed as a normal probability density about the overall process mean (assumed to be zero in this case) with process variance σ^2 common for all t. Assumption 2 indicates that the values of θ at successive times are not independent and their joint density is bivariate normal with correlation coefficient ρ. The observation and process model for the autoregressive process can be expressed as:

$$\text{OBSERVATION:} \quad x_t = \theta_t + e_t \qquad e_t \sim N(0,\sigma_x^2) \tag{34}$$

$$\text{PROCESS:} \quad \theta_t = \rho\theta_{t-1} + w_t \qquad w_t \sim N(0,\sigma^2(1-\rho^2)) \tag{35}$$

This model is useful when there appears to be no underlying trend either up or down in the process and the quality of successive lots appears to be highly correlated. From the above model

$$f(\theta_t|\theta_{t-1}) \sim N(\rho\theta_{t-1},\sigma^2(1-\rho^2))$$

C. PERIODIC MODEL

If the production process behaves in periodic fashion over some internal T then a periodic model is appropriate. Let $\theta(t)$ represent the periodic function which describes the production process and let the average value of the periodic function be zero (i.e., $\frac{1}{T}\int_0^T \theta(t)dt = 0$). Then the periodic function $\theta(t)$ can be approximated by the Fourier series as

$$\theta(t) = \sum_{k=1}^{n} (a_k \text{COS} k\omega_o t + b_k \text{SIN} k\omega_o t) \quad , \quad \omega_o = \frac{2\pi}{T}$$

where the coefficients a_k and b_k are unknown and subject to disturbances. The value of n being large enough to make the approximation valid.

Let $\theta^T = (a_1 a_2 \ldots A_n b_1 b_2 \ldots b_n)$

and $B^T = (\cos\omega_o t \cos 2\omega_o t \ldots \cos n\omega_o t \sin\omega_o t \ldots \sin n\omega_o t)$

Then the observation and process models are:

$$\text{OBSERVATION:} \quad x_t = B_t^T \theta_t + e_t \quad e_t \sim N(0, \sigma_x^2) \tag{36}$$

$$\text{PROCESS:} \quad \theta_t = \theta_{t-1} + w_t \quad w_t \sim N(\underline{0}, \Sigma) \tag{37}$$

where $\underline{0}$ is a 2nx1 vector of zeros and $\Sigma = E[w_t w_t^T]$.

D. GENERAL LINEAR MODEL

The above models represent special cases of the generalized linear model incorporating special features to reflect particular characteristics of the production process. The generalized linear model is defined as follows:

Let $\underline{x}_t$ = nx1 vector of observations at time t

θ_t = px1 vector of process parameters at time t

A_1 = nxp matrix characterizing the observation

A_2 = pxp matrix characterizing the process

e_t = nx1 vector of observation noise at time t

ω_t = px1 vector of process noise at time t

$E[e_t] = E[w_t] = \underline{0}$

VAR $(e_t) = E[e_t e_t^T] = C_1$

VAR $(w_t) = E[w_t w_t^T] = C_2$

Then OBSERVATION $\underline{x}_t = A_1\theta_t + e_t \quad e_t \sim N(\underline{0}, C_1)$ (38)

PROCESS $\theta_t = A_2\theta_{t-1} + w_t \quad w_t \sim N(\underline{0}, C_2)$ (39)

is the generalized linear model.

In order to implement the Bayesian decision procedure various probability densities are required to determine the risks of alternative actions. From section II it can be seen that three densities must be obtained in the course of the procedure. In order to obtain the risk of immediate decision without sampling the prior distribution of θ, $f(\theta)$ must be obtained. To obtain the risk of sampling the prior distribution of $\underline{x}$, $f(\underline{x})$ and the conditional distribution of θ given a sample $\underline{x}$, $f(\theta|\underline{x})$ must be obtained. In order to determine the optimal decision after sampling $f(\theta|\underline{x})$ is required. Thus to implement the decision procedure three densities must be obtained at each step in the process. If the observation and

production process can be modeled by the generalized linear model of (38) and (39) the required densities can be obtained as follows [4].

Let: $f(\theta_t)$ be the density of θ at time t prior to sampling

$f(\theta_t|\underline{x})$ be the posterior of θ at time t based on the sample $\underline{x}$

$f(\underline{x}_t)$ be the sample distribution prior to sampling

Also let the distribution of θ at t-1 be $N(\mu,\Sigma)$

Then from (38) and (39)

$$f(\theta_t) \sim N(A_2\mu, A_2\Sigma A_2^T + C_2) \tag{40}$$

$$f(\underline{x}_t) \sim N(A_1A_2\mu,\ A_1(A_2\Sigma A_2^T + C_2)A_1^T + C_1) \tag{41}$$

$f(\theta_t|x_t)$ is obtained as follows (where the matrices for which inverses are needed are assumed non-singular). From Bayes theorem $f(\theta_t|\underline{x}_t) \propto f(\underline{x}|\theta_t)\ f(\theta_t)$

From the model $f(\underline{x}_t|\theta_t) \sim N(A_1\theta_t,\ C_1)$

Thus $f(\theta_t|\underline{x}_t) \propto e^{-\frac{1}{2}Q}$

where $Q = (\underline{x} - A_1\theta)^T C_1^{-1}(\underline{x}-A_1\theta) +$

$$(\theta-A_2\mu)^T(A_2\Sigma A_2^T + C_2)^{-1}(\theta-A_2\mu)$$

the t subscripts being deleted for convenience, collecting the θ terms yields:

$$Q = \theta^T(\Delta+A_1^TC_1^{-1}A_1)\theta - 2(\underline{x}^TC_1^{-1}A_1 + \mu^TA_2^T\Delta)\theta$$
$$+ \underline{x}^TC_1^{-1}\underline{x} + \mu^TA_2^T\Delta A_2\mu$$

where $\Delta = (A_2\Sigma A_2^T + C_2)^{-1}$

completing the square results in:

$$Q = (\theta_1-Dd)^T D^{-1}(\theta_1-Dd) + [\underline{x}^TC_1^{-1}\underline{x} + \mu^TA_2^T\Delta A_2\mu - d^TD^Td]$$

Thus $f(\theta_t|\underline{x}_t) \sim N(Dd,D)$ (42)

where $D^{-1} = (A_2\Sigma A_2^T + C_2)^{-1} + A_1^TC_1^{-1}A_1$

$$d = A_1^TC_1^{-1}\underline{x} + (A_2\Sigma A_2^T + C_2)^{-1}A_2\mu$$

Thus with the aid of the generalized linear model the required densities for complex multidimensional production processes can be evaluated in a straightforward manner using (40), (41) and (42).

Equations (40) and (41) represent the one step ahead predictive distributions of θ and x based on the prior at t-1. As discussed in section II, if no sampling is performed at some t then the predictive distribution for θ at t+1 based on the prior at t-1 is required. In general, a method is needed to obtain the k^{TH} step ahead predictive distributions of θ and x.

Let $f(\theta_t) \sim N(\mu_t,\Sigma_t)$ then the k^{TH} step ahead distribution $f(\theta_{t+k}) \sim N(\mu_{t+k}, \Sigma_{t+k})$ can be obtained recursively from the linear model as follows:

$$\mu_{t+k} = A_2\mu_{t+k-1} \quad , \quad k = 0,1,2, \ldots \tag{43}$$

$$\Sigma_{t+k} = A_2\Sigma_{t+k-1}A_2^T + C_2, \quad k=0,1,2, \ldots \tag{44}$$

Let $f(\underline{x}_{t+k}) \sim N(m_{t+k},C_{t+k})$ be the k^{TH} step ahead predictive distribution for the sample $\underline{x}_{t+k}$ then from (43) and (44) and the linear model, m_{t+k} and C_{t+k} can be determined recursively by:

$$m_{t+k} = A_1\mu_{t+k} \quad , \quad k = 0,1,2, \ldots \tag{45}$$

$$C_{t+k} = A_1\Sigma_{t+k}A_1^T + C_1, \quad k = 0,1,2, \ldots \tag{46}$$

For an example of the use of the linear model and the risk calculations the reader is referred to Appendix A. As an interesting aside, the recursion relationships developed in (42) thru (46) are identical to the results obtained using Kalman filtering.

V. INCORPORATING SUBJECTIVE INFORMATION

One of the primary advantages of the Bayesian decision procedure is its capacity to incorporate subjective information into the decision process. Subjective information can be incorporated into the decision process by either of two routes, either by revisions to the process model or by altering the prior distribution of θ. The method chosen depending on which more accurately reflects the subjective information. Examples of how subjective information may be used will be discussed with respect to the generalized linear model which is repeated here for reference

$$\text{OBSERVATION:} \quad \underline{x}_t = A_1\theta_t + e_t, \quad e_t \sim N(\underline{0}, C_1)$$

$$\text{PROCESS:} \quad \theta_t = A_2\theta_{t-1} + w_t, \quad w_t \sim N(\underline{0}, C_2)$$

As an example of the use of subjective information, suppose that the autoregressive model of section IV is being used to model the production process and production appears to be fairly stable (i.e., no trends). You are informed that starting with the next production lot three engineering changes will be incorporated into the units. It has been your experience that whenever more than one engineering change is incorporated that the production quality is momentarily reduced and then increases with successive lots as the new procedures are learned and the inspectors gain experience.

How might this subjective information be incorporated into the decision process? One approach would be to adjust the prior for the first change lot by lowering the mean to reflect the anticipated decrease in quality and increasing the variance to reflect the associated uncertainty as to the actual process value. Increasing the variance will have the effect of weighting the new sample data more heavily in determining the posterior. After adjusting the prior to reflect the anticipated decrease in quality the autoregressive model might be replaced by the linear trend model to reflect the anticipated increase of quality with successive lots as a result of the learning process. The rate of increase δ_t in the linear model could be changed for each lot to provide a linear approximation to the anticipated learning curve. As the process quality returned to its original level the autoregressive model would again be used.

This example illustrates two important features of the Bayesian decision procedure and the use of the linear model. First, when using a linear model it is not necessary that A_1, A_2, C_1 and C_2 be constant for all t only that they be known at time t and thus the parameters of the linear model are free to change as required by the process being modeled. The second feature of the Bayesian procedure illustrated in the example is adaptability. By changing the model structure or the prior to reflect uncertainty in the process, the information requirements (sample data) of the procedure

adapt to reflect these changes. In order to maintain the same risk more samples will be required if the variance of the prior is increased to reflect uncertainty. Thus unlike traditional quality control procedures where the same sampling and decision procedure is used the Bayesian method can adapt to reflect the changing requirements of the production process. As another example of adaptability consider the autoregressive model. Because of the high correlation from one lot to the next the method reduces the sampling required when quality is either very good or very poor, thus taking advantage of the natural excursions of the output quality.

The adaptability feature of the Bayesian method also has another interesting property. It indicates where, when and the quantity of quality control resources to be used. This is especially important when trying to maximize the effectiveness of the quality control function on fixed or limited resources (i.e., labor, test facilities, etc.).

APPENDIX A

A BAYESIAN DECISION EXAMPLE

The following example is provided to illustrate how the Bayesian method is applied and the required risks are obtained. It is assumed that an autoregressive model is used to represent the production process and that simultaneous sampling is used.

Let: OBSERVATION MODEL: $x_t = \theta_t + e_t$, $e_t \sim N(e, \sigma_x^2)$

PROCESS MODEL: $\theta_t = \rho\theta_{t-1} + w_t$

$w_t \sim N(0, \sigma^2(1-\rho^2))$

and the loss function is:

$$L(\theta, d_1) = \exp[-(\theta-\theta_0)] , \quad -\infty \leq \theta \leq \infty$$

$$L(\theta, d_2) = \exp[-(\theta_0-\theta)], \quad -\infty \leq \theta \leq \infty$$

where : θ_0 is a known constant.

Also let $f(\theta_{t-1}) \sim N(\mu_{t-1}, \sigma^2_{t-1})$ be the density of θ at time t-1.

From section IV the correspondence between the autoregressive model and the generalized linear model is: $A_1 = 1$, $A_2 = \rho$, $C_1 = \sigma_x^2$, and $C_2 = \sigma^2(1-\rho^2)$. Thus by equation (40) and (41) making the above substitutions:

$$f(\theta_t) \sim N(\rho\mu_{t-1},\ \rho^2\sigma_{t-1}{}^2 + (1-\rho^2)\sigma^2)$$

$$f(x) \sim N(\rho\mu_{t-1},\ \rho^2\sigma_{t-1}{}^2 + (1-\rho^2)\sigma^2 + \sigma_x{}^2)$$

let $\mu_t = \rho\mu_{t-1}$ and $\sigma_t{}^2 = \rho^2\sigma_{t-1}{}^2 + (1-\rho^2)\sigma^2$

then $f(\theta_t) \sim N(\mu_t, \sigma_t{}^2)$ (1)

$$f(x) \sim N(\mu_t, \sigma_t{}^2 + \sigma_x{}^2) \quad (2)$$

To determine the risk of making an immediate decision from equation (10)

$$\rho(d_1) = \int_{-\infty}^{\infty} L(\theta, d_1)\ f(\theta_t)d\theta_t = \exp[\frac{\sigma_t{}^2}{2} - (\mu_t - \theta_0)] \quad (3)$$

and

$$\rho(d_2) = \int_{-\infty}^{\infty} L(\theta, d_2)\ f(\theta_t)d\theta_t = \exp[\frac{\sigma_t{}^2}{2} + (\mu_t - \theta_0)] \quad (4)$$

Equations (3) and (4) are plotted in figure A-1 as a function of the mean μ_t.

Since $\rho(d^*)$, the risk of the optimal decision, is equal to the minimum risk from figure A-1 it is observed that:

$$\rho(d^*) = \begin{cases} \rho(d_2) & \mu_t < \theta_0 \\ \rho(d_1) & \mu_t \geq \theta_0 \end{cases} \quad (5)$$

which implies the following decision rule:

$$d^* = \begin{cases} d_2 \text{ if } & \mu_t < \theta_0 \\ d_1 \text{ if } & \mu_t \geq \theta_0 \end{cases}$$

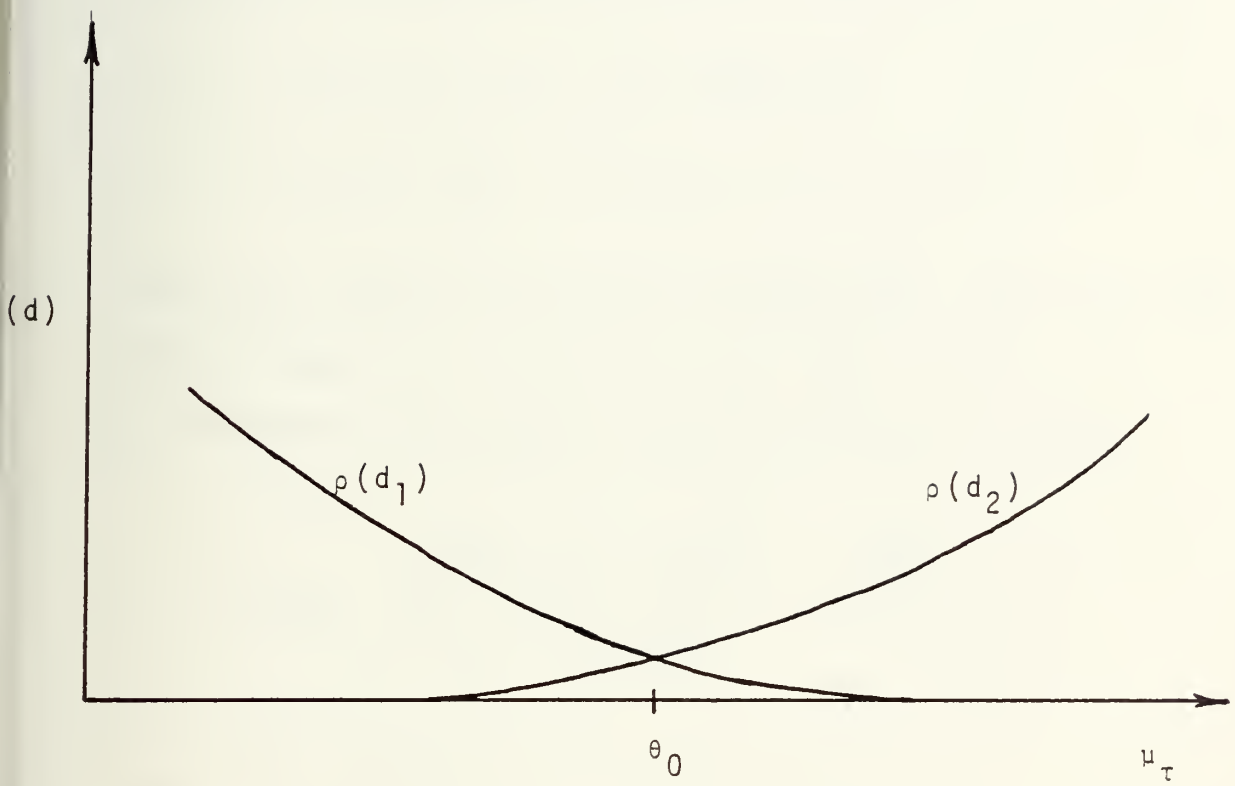

Figure A-1. Decision Risk.

By symmetry the risk of decision d* is

$$\rho(d^*) = \exp\left[\frac{\sigma_t^2}{2} - |\mu - \theta_0|\right]$$

In order to determine the risk of sampling and then making a decision d* the posterior $f(\theta_t|\underline{x})$ must be obtained. The process of testing a sample of size n parameterized by a fixed but unknown value θ_t can be viewed as:

$$x_i = \theta_i + e_i \qquad e \sim N(0, \sigma_x^2)$$

$$\theta_i = \theta_{i-1} \qquad i = 1,2, \ldots, n$$

where the samples are assumed independent. Applying the linear model: $A_1 = 1$, $A_2 = 1$, $C_1 = \sigma_x^2$, and $C_2 = 0$. By repeated application of equation (40) one obtains:

$$f(\theta_t|\underline{x}) \sim N\left(\frac{\bar{x}_n \sigma_t^2 + \mu_t \frac{\sigma_x^2}{n}}{\sigma_t^2 + \frac{\sigma_x^2}{n}}, \frac{\sigma_t^2 \frac{\sigma_x^2}{n}}{\sigma_t^2 + \frac{\sigma_x^2}{n}}\right)$$

where: $\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$

From section II the risk of obtaining an additional sample and then choosing d* is:

$$\rho(\underline{x}) = \int_{\underline{X}} \{\inf_D \int_\Theta \rho(\theta,d) f(\theta|\underline{x}) d\theta\} f(\underline{x}) d\underline{x} \qquad (6)$$

the expression in brackets is $\rho(d^*|\underline{x})$ that is, the risk of d^* given a sample result $\underline{x}$. $\rho(d_1|\underline{x})$ and $\rho(d_2|\underline{x})$ may be obtained by substituting $f(\theta_t|\underline{x})$ for $f(\theta_t)$ in equations (3) and (4). This substitution yields:

$$\rho(d_1|\underline{x}) = \exp[\tfrac{1}{2}\alpha^2 - (\beta-\theta_0)]$$

$$\rho(d_2|\underline{x}) = \exp[\tfrac{1}{2}\alpha^2 + (\beta-\theta_0)]$$

where

$$\alpha = \frac{\sigma_t^2 \frac{\sigma_x^2}{n}}{\sigma_t^2 + \frac{\sigma_x^2}{n}} \quad , \quad \beta = \frac{\bar{x}_n \sigma_t^2 + \mu_t \frac{\sigma_x^2}{n}}{\sigma_t^2 + \frac{\sigma_x^2}{n}}$$

From equation (5)

$$\rho(d^*|\underline{x}) = \begin{cases} \rho(d_2|\underline{x}) , & \beta < \theta_0 \\ \rho(d_1|\underline{x}) , & \beta \geq \theta_0 \end{cases}$$

Equation 6 becomes:

$$\rho(\underline{x}) = \int_{-\infty}^{\bar{x}_c} \rho(d_2|\bar{x}_n)\, f(\bar{x}_n)\, d\bar{x}_n + \int_{\bar{x}_c}^{\infty} \rho(d_1|\bar{x}_n)\, f(\bar{x}_n) d\bar{x}_n \qquad (7)$$

where: $\bar{x}_c = \frac{\sigma_x^2(\theta_0-\mu_t)}{n\sigma_t^2} + \theta_0$ and $f(\bar{x}_n) \sim N(\mu, \sigma_t^2+\frac{\sigma_x^2}{n})$

Solving (7) yields

$$\rho(\underline{x}) = \rho(d_1)\, \bar{\Phi}\left(\frac{\bar{x}_c+\sigma_t^2-\mu}{\sqrt{\sigma_t^2 + \frac{\sigma_x^2}{n}}}\right) + \rho(d_2)\, \Phi\left(\frac{\bar{x}_c - \sigma_t^2 - \mu}{\sqrt{\sigma_t^2 + \frac{\sigma_x^2}{n}}}\right) \qquad (8)$$

where: $\rho(d_1)$ and $\rho(d_2)$ are as defined in (3) and (4)

$$\Phi(\cdot) = P\{z<\cdot\}, \quad z \sim N(0,1)$$

$$\bar{\Phi}(\cdot) = 1-\Phi(\cdot)$$

Based on equations (5) and (8) the usual decision rule is applied. If $\rho(d^*) > \rho(\underline{x}) + E[C(x)]$ take another sample otherwise make decision d^*.

To determine the mean and variance of the k^{TH} step ahead predictive distribution μ_{t+k} and σ^2_{t+k} equations (43) and (44) are used recursively to obtain the following results.

$$\mu_{t+k} = \rho^k \mu_t \tag{9}$$

$$\sigma^2_{t+k} = \rho^{2k}\sigma_t^2 + \sigma^2(1-\rho^{2k}) \tag{10}$$

The predictive distribution for k^{TH} step ahead sample parameterized by m_{t+k} and C_{t+k} by (45) and (46) are:

$$m_{t+k} = \mu_{t+k}$$

$$C_{t+k} = \sigma^2_{t+k} + \sigma_x^2$$

In order to examine the behavior of the posterior of θ, $f(\theta|\underline{x})$ as the sample size is increased observe that

$$\lim_{n\to\infty} f(\theta_t|\underline{x}) = \lim_{n\to\infty} N\left(\frac{\bar{x}_n\sigma_t^2 + \mu\frac{\sigma_x^2}{n}}{\sigma_t^2 + \frac{\sigma_x^2}{n}}, \frac{\sigma_x^2}{n+\frac{\sigma_x^2}{\sigma_t^2}}\right)$$

Thus $\lim_{n\to\infty} f(\theta|\underline{x}) \to N(\theta_t, 0)$

Since $\bar{x}_n$ is a consistent estimator of θ_t. This implies that as n increases the knowledge of θ_t as represented by $f(\theta_t|\underline{x})$ becomes "perfect" in the sense that the variance approaches zero and θ_t is then known. The variance decreases approximately as $\frac{1}{n}$ for $n >> \frac{\sigma_x^2}{\sigma_t^2}$.

The k^{TH} step ahead prediction distribution, $f(\theta_{t+k})$ of (9) and (10) represents the uncertainty in θ_{t+k} based on information up to and including time t. The $\lim_{k\to\infty} f(\theta_{t+k}) \to N(0,\sigma^2)$ which is the distribution of the process before any information is obtained. Thus as k increases the information obtained at time t loses its "value" in predicting the location of the process at t+k. The rate at which previous knowledge is discounted is a function of ρ the correlation between θ_{t+1} and θ_t which in this example was assumed constant for all t.

REFERENCES

1. DeGroot, M. H., Optimal Statistical Decisions, McGraw Hill, 1970.

2. Chernoff, H. and Moses, L., Elementary Decision Theory, Wiley, 1959.

3. Hamburg, M., "Bayesian Decision Theory and Statistical Quality Control," Industrial Quality Control, vol. 19, No. 6, December 1962.

4. Lindley, D. V. and Smith, A. F. N., "Bayes Estimates for the Linear Model," J.R. Statistical Society, B, 34, No. 1, 1972, pp. 1-18.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, VA 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, CA 93940	2
3.	Department Chairman, Code 55 Department of Operations Research Naval Postgraduate School Monterey, CA 93940	1
4.	Assoc. Professor B. O. Shubert, Code 55Sy Department of Operations Research Naval Postgraduate School Monterey, CA 93940	1
5.	Assoc. Professor Glen Lindsay, Code 55Ls Department of Operations Research Naval Postgraduate School Monterey, CA 93940	1
6.	R. D. Desposato, Code 1141 Head Pate Engineering Branch Pacific Missile Test Center Point Mugu, CA 93042	1
7.	James R. Anderson 1775 Shannon Ave. Ventura, CA 93003	5

Thesis
.A4543
c.1

170388

Anderson
Bayesian decision theory in variables sampling for quality control.